

PATENT

MS303968.01/MSFTP444US

CERTIFICATE OF FACSIMILE TRANSMISSION

I hereby certify that this correspondence (along with any paper referred to as being attached or enclosed) is being faxed to **571-273-8300** on the date shown below to **Mail Stop Appeal Brief - Patents**, Commissioner for Patents, P.O. Box 1450, Alexandria, Virginia 22313-1450.

Date: February 5, 2009

/Luke Clossman/
Luke Clossman

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re patent application of:

Applicants(s): Eric D. Brill, *et al.*

Serial No: 10/600,797

Filing Date: June 20, 2003

Examiner: Michael J. Hicks

Art Unit: 2165

Title: SYSTEMS AND METHODS TO TUNE A GENERAL-PURPOSE SEARCH ENGINE FOR A SEARCH ENTRY POINT

**Mail Stop Appeal Brief-Patents
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450**

APPEAL BRIEF

Dear Sir:

Appellants submits this brief in connection with an appeal of the above-identified patent application. A credit card payment form is filed concurrently herewith in connection with all fees due regarding this appeal brief. In the event any additional fees may be due and/or are not covered by the credit card, the Commissioner is authorized to charge such fees to Deposit Account No. 50-1063 [MSFTP444US].

I. Real Party in Interest (37 C.F.R. §41.37(c)(1)(i))

The real party in interest in the present appeal is Microsoft Corporation, the assignee of the present application.

II. Related Appeals and Interferences (37 C.F.R. §41.37(c)(1)(ii))

Appellants, appellants' legal representative, and/or the assignee of the present application are not aware of any appeals or interferences which may be related to, will directly affect, or be directly affected by or have a bearing on the Board's decision in the pending appeal.

III. Status of Claims (37 C.F.R. §41.37(c)(1)(iii))

Claim 41 has been cancelled. Claims 1-40, 42, and 43 stand rejected by the Examiner. The rejection of claims 1-40, 42, and 43 is being appealed.

IV. Status of Amendments (37 C.F.R. §41.37(c)(1)(iv))

The Examiner has entered amendments submitted after the Final Office Action. (*See* Communication from Examiner, Advisory Action dated October 23, 2008, continuation of item 13).

V. Summary of Claimed Subject Matter (37 C.F.R. §41.37(c)(1)(v))**A. Independent Claim 1**

Independent claim 1 recites, "A system that refines a general-purpose search engine, comprising: a component that identifies an entry point that includes a link utilized to access the general-purpose search engine; and a tuning component that receives search query results of the general-purpose search engine and filters the search results based at least on criteria associated with the entry point through which the general-purpose search engine was accessed, the criteria comprises at least a first set of data categorized as relevant to a user's context and a second set of data categorized as non-relevant to the user's context, wherein user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data, the first and second sets of data persisted to a

computer-readable storage medium.” (*See e.g.*, FIG. 1; paragraphs [0029], [0030], [0042], and [0045]).

B. Independent Claim 13

Independent claim 13 recites, “A system that tunes a general-purpose search engine, comprising: a filter component that receives search query results of a general-purpose search engine and parses relevant and non-relevant results based on training data associated with the entry point that provides a link employed to traverse to the general-purpose search engine, the training data comprises a first set of data categorized as relevant to a search context of a user for the entry point and a second set of data categorized as non-relevant to the search context of the user; and a ranking component that sorts the filtered results in accordance with the training data for presentation to a user, wherein a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be automatically added to the second set of data, the first and second sets of data persisted to a computer-readable storage medium.” (*See e.g.*, FIG. 3; paragraphs [0039], [0041], and [0045]).

C. Independent Claim 22

Independent claim 22 recites, “A method to filter and rank general-purpose search engine results based on criteria associated with an entry point, comprising: executing a query search with the general-purpose search engine accessed through a link associated with the entry point; filtering the general-purpose search engine results by tuning the general-purpose search engine based on a set of training data associated with the entry point employed to access the general purpose search engine; ranking the filtered general-purpose search engine results; automatically storing a first query result selected by a user in a first data set categorized as relevant; automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon selection of the first query result; and including the first data set and second data set in the set of training data associated with the entry point employed to access the general purpose search engine.” (*See e.g.*, FIG. 7; paragraphs [0010], [0011], [0045], and [0067]).

D. Independent Claim 29

Independent claim 29 recites, “A method to customize a general-purpose search engine to improve context search query results, comprising: tuning a general-purpose search engine for an entry point by employing a method further comprising: providing a first set of data categorized as relevant that is used by a component to discern query results relevant to a search context of a user employing the entry point, the entry point provides a link employed to access the general-purpose search engine; providing a second set of data categorized as non-relevant that is used by the component to discern query results unrelated to the search context, the first set of data and the second set of data are manually provided; determining whether a query result is relevant or non-relevant to the search context based on the first set of relevant data and the second set of non-relevant data, each query result is compared with both the first set of data and second set of data to determine the relevance of the query result; executing a search query with the general purpose search engine to obtain a ranked list of query results; selecting a link associated with a query result from the list; automatically adding the selected query result to the first set of data; and automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of the selected query result.” (*See e.g.*, Figure 3; paragraphs [0042], [0043], and [0045]).

E. Independent Claim 34

Independent claim 34 recites, “A method to automatically customize a general-purpose search engine for an entry point, comprising: identifying the entry point; executing a query search *via* the entry point that includes a link employed to route to the general-purpose search engine; recording a first query result from a ranked list of query results returned from the executed query as relevant when a user views the document associated with the first query result; recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when the first result is selected for viewing by the user; and providing the recorded results to automatically train the filter for the entry point, in order to discriminate between results relevant to a search context of the user for the entry point and results non-relevant to the search context.” (*See e.g.*, paragraphs [0038], [0066], [0045], [0046]).

F. Independent Claim 42

Independent claim 42 recites, “A computer readable storage medium storing computer executable components that tunes a general-purpose search engine to improve context search query results, comprising: a component that receives search query results of a general-purpose search engine and filters the results based on training data sets associated with the search context of a user depending on the entry point that provides a link utilized to arrive at the general-purpose search engine, the training data sets include at least a first category of data explicitly defined to be relevant to the search context and a second category of data explicitly defined to be non-relevant to the search context; and a component that ranks the filtered general-purpose search engine results according to the similarity of the search engine results to the training data sets, wherein selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data.” (*See e.g.*, FIG. 1; paragraphs [0029], [0030], [0042], and [0045]).

G. Independent Claim 43

Independent claim 43 recites, “A system that receives, filters and ranks general-purpose search engine results, comprising: means for filtering general-purpose search engine results by determining whether a query result is relevant to a search context of a group of users, the search context is associated with an entry point that includes a link employed to navigate to the general-purpose search engine, the search context further having an associated first set of training data categorized as relevant to the context and an associated second set of training data categorized as non-relevant to the context; and means for ranking the filtered general-purpose search engine results based on a relevance of the general-purpose search engine results to the search context of the group of users and the entry point as determined by a comparison of the search engine results with the first and second sets of training data, wherein a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of training data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of training data, the first and second sets of training data stored on a computer-readable storage medium.” (*See e.g.*, FIG. 3; paragraphs [0039], [0041], and [0045]).

VI. Grounds of Rejection to be Reviewed (37 C.F.R. §41.37(c)(1)(vi))

A. Whether claims 1-6, 8-16, 18-22, 29-40, and 42-43 are unpatentable under 35 U.S.C. §102(a) over of Joachims ("Optimizing Search Engines Using Clickthrough Data," Proceedings of the Eighth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, Pages 133-142, 2002, ACM).

B. Whether claims 7, 17, and 23-28 are unpatentable under 35 U.S.C. §103(a) over Joachims ("Optimizing Search Engines Using Clickthrough Data," Proceedings of the Eighth ACM SIGKDD International Conference of Knowledge Discovery and Data Mining, Pages 133-142, 2002, ACM) in view of Pazzani, *et al.* ("Learning and Revising User Profiles: The Identification of Interesting Web Sites," Machine Learning 27, Pages 313-331, 1997, Kluwer Academic Publishers).

VII. Argument (37 C.F.R. §41.37(c)(1)(vii))

A. **Rejection of Claims 1-6, 8-16, 18-22, 29-40, and 42-43 Under 35 U.S.C. §102(a)**

Claims 1-6, 8-16, 18-22, 29-40, and 42-43 stand rejected under 35 U.S.C. §102(a) as being anticipated by Joachims. Reversal of this rejection is respectfully requested for at least the following reasons. Joachims fails to teach or suggest each and every limitation as recited in the subject claims.

For a prior art reference to anticipate, 35 U.S.C. §102 requires that "each and every element as set forth in the claim is found, either expressly or inherently described, in a single prior art reference." *In re Robertson*, 169 F.3d 743, 745, 49 USPQ2d 1949, 1950 (Fed. Cir. 1999) (*quoting Verdegaal Bros., Inc. v. Union Oil Co.*, 814 F.2d 628, 631, 2 USPQ2d 1051, 1053 (Fed. Cir. 1987)).

The subject claims relate to refinement of search query results from a general-purpose search engine based in part on the entry point through which the search engine was accessed. When the search engine is accessed *via* an entry point and a search query is executed, the search

query results obtained by the search engine can be passed to a tuning component associated with the entry point (*See e.g.*, paragraph [0029]). The tuning component can filter and rank the search query results according to a statistical analysis that utilizes two distinct sets of training data associated with the entry point and the context of the search: a first set of data expressly defined as relevant to the search context and a second data set expressly defined as non-relevant to the search context (*See e.g.*, paragraphs [0010], [0042]). The first and second data set can be automatically trained based on observance of user selections. Specifically, when a user selects one of the filtered search query results presented by the tuning component, that result can be automatically recorded by the training component as relevant to the search context and added to the first data set, while results that had been ranked higher than the selected result can be automatically added to the category of non-relevant data (*See e.g.*, paragraphs [0011], [0045]). In particular, independent claim 1 recites, *user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data, the first and second sets of data persisted to a computer-readable storage medium.*

Contrary to the Examiner's assertions, Joachims does not disclose such a technique for maintaining sets of training data. Joachims relates to optimization of search result rankings through analysis of recorded clickthrough data. During a search session, clickthrough data is recorded in a logfile for analysis. This clickthrough data is recorded in the form of triplets consisting of a user query (q), the ranking of results (r) presented to the user in response to the query, and the links subsequently selected by the user (c). Once recorded, the clickthrough data is used to train a retrieval function intended to improve search result rankings based on ranking preferences inferred from the logfile. However, the cited reference does not disclose that *selection of a query result can cause non-selected but higher ranked results to be added to a "non-relevant" training data set*, since the training data employed in Joachims' approach does not include both a first data set categorized as "relevant" *and a second data set categorized as "non-relevant."* Indeed, Joachim expressly declares that such a "binary classification" runs counter to the approach explored in that reference. The first paragraph of Section 4, for example, states, "Most work on machine learning in information retrieval does not consider the formulation of above, but simplifies the task to a binary classification problem with the two

classes ‘relevant’ and ‘non-relevant’. Such a simplification has several drawbacks...Therefore, the following algorithm directly addresses taking an empirical risk minimization approach.” Moreover, the same paragraph concedes that such a data set of non-relevant results *cannot be derived from the clickthrough training data* utilized by Joachims, stating that “such absolute relevance judgments [that is, classifying a result to be non-relevant] *cannot be extracted from clickthrough data.*” Clearly, Joachims expressly foregoes the use of “relevant” and “non-relevant” training data sets, and instead attempts to infer relative degrees of relevance from the aforementioned clickthrough triplet data. Since Joachims admittedly does not utilize a training data set of “non-relevant” results, it cannot be said that Joachims in any way teaches or suggests that selection of a search result can cause non-selected but higher ranked results to be added to such a list.

Contending that Joachims discloses these technique for populating a training set of non-relevant data, the Examiner notes that selection of a link from a search result set generated using Joachims’ approach causes the clickthrough data for that selection to be recorded, and that this recorded data includes those results not selected by the user. However, the cited reference teaches that this clickthrough data is recorded in the triplet format discussed above (Section 5.1, second paragraph: “The clicks of the user are recorded using the proxy system described in Section 2.1,” wherein Section 2.1 discloses recording the *clickthrough data triplets*). As already noted, this data includes the query (q), the ranking (r) presented to the user, and the links selected by the user (c). Presumably, the results not selected by the user are recorded in the ranking (r) information, together with the results that *were* selected by the user. Such a record does not constitute a separate set of data categorized as non-relevant, since this data is only recorded as part of the *total* ranking of results presented to the user. By contrast, the subject claims teach that selection of a search result can cause *both distinct lists* (relevant and non-relevant) to be populated with new training data *via* the techniques disclosed in independent claim 1. Joachims, which expressly foregoes such a binary classification of training data, in no way teaches or suggests these techniques.

Independent claim 13 recites, *a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be*

automatically added to the second set of data. As discussed *supra*, Joachims fails to teach or suggest this technique for populating a set of relevant and a set of non-relevant training data.

Independent claim 22 recites, *ranking the filtered general-purpose search engine results; automatically storing a first query result selected by a user in a first data set categorized as relevant; automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon selection of the first query result.* These aspects are not disclosed in Joachims, as already discussed.

Independent claim 29 recites, *selecting a link associated with a query result from the list; automatically adding the selected query result to the first set of data; and automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of the selected query result.* Joachims is silent regarding these aspects, as discussed above.

Moreover, independent claim 34 recites, *recording a first query result from a ranked list of query results returned from the executed query as relevant when a user views the document associated with the first query result; recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when the first result is selected for viewing by the user,* and as already discussed, the cited reference is silent regarding these aspects.

Independent claim 42 recites, *selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data.* The cited reference does not disclose these techniques for collecting training data, as noted above.

Independent claim 43 recites, *a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of training data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of training data.* As discussed *supra*, Joachims fails to teach or suggest these features.

In view of at least the foregoing, it is respectfully submitted that Joachims does not teach or suggest each and every feature of independent claims 1, 13, 22, 29, 34, 42, and 43 (and all claims depending there from), and as such fails to anticipate the claimed subject matter. It is

therefore requested that this rejection be reversed.

B. Rejection of Claims 7, 17, and 23-28 Under 35 U.S.C. §103(a)

Claims 7, 17, and 23-28 stand rejected under 35 U.S.C. §103(a) as being unpatentable over Joachims in view of Pazzani, *et al.* Reversal of this rejection is respectfully requested for at least the following reasons. Joachims and Pazzani, *et al.*, individually or in combination, do not teach or suggest all features set forth in the subject claims.

Claim 7 depends from independent claim 1, claim 17 depends from independent claim 13, and claims 23-28 depend from independent claim 22. As discussed in the previous section of the Reply with respect to those independent claims, Joachims does not teach or suggest that selection of a search result can cause the selected result to be added to a training data set of relevant results, while causing non-selected but higher ranked search results to be added to a data set of non-relevant results. Pazzani, *et al.* is also silent regarding this manner of collecting training data. Pazzani, *et al.* presents an overview of an intelligent agent called Syskill & Webert, which is used to develop and refine user profiles that infer websites of interest to the associated user. These user profiles can be revised and updated in response to feedback from the associated user regarding which websites are of interest and which are not, and these updated profiles can be used to predict which websites will be of most interest to the user. Although Pazzani, *et al.* teaches that algorithms used to update the user profiles employ a set of positive examples (*e.g.* websites of interest to the user) and negative examples (*e.g.* websites the user is not interested in), the cited reference indicates that these examples must be *explicitly selected* by the user in both cases. Section 2.1, paragraph 3, for example, explains that pages are rated by a user as being “hot” (interesting), or “cold” (not interesting), and these ratings are used to train the algorithm that refines the user profiles. Hence, websites that a user finds uninteresting for a given search session must be explicitly selected by the user for inclusion in the negative examples. The subject claims, by contrast, disclose that query results ranked higher than a selected result can automatically be included in the set of non-relevant training data without the need to visit or rank the non-relevant sites, thereby removing the burden of explicitly indicating uninteresting websites from the user. Pazzani, *et al.* does not disclose such an automated technique for collecting training data.

In view of at least the foregoing, it is respectfully submitted that Joachims, alone or in combination with Pazzini, *et al.*, does not teach or suggest all aspects set forth in independent claims 1, 13, and 22 (and all claims depending there from), and as such fails to make obvious the claimed subject matter. It is therefore requested that this rejection be reversed.

F. Conclusion

For at least the above reasons, the claims currently under consideration are believed to be patentable over the cited references. Accordingly, it is respectfully requested that the rejections of claims 1-40, 42, and 43 be reversed.

If any additional fees are due in connection with this document, the Commissioner is authorized to charge those fees to Deposit Account No. 50-1063 [MSFTP444US].

Respectfully submitted,
AMIN & TUROCY, LLP

Himanshu S. Amin
Reg. No. 40,894

AMIN & TUROCY, LLP
24th Floor, National City Center
1900 East 9th Street
Telephone: (216) 696-8730
Facsimile: (216) 696-8731

VIII. Claims Appendix (37 C.F.R. §41.37(c)(1)(viii))

1. A system that refines a general-purpose search engine, comprising:
a component that identifies an entry point that includes a link utilized to access the general-purpose search engine; and
a tuning component that receives search query results of the general-purpose search engine and filters the search results based at least on criteria associated with the entry point through which the general-purpose search engine was accessed, the criteria comprises at least a first set of data categorized as relevant to a user's context and a second set of data categorized as non-relevant to the user's context, wherein user selection of a query result from a ranked list of the query results causes the selected result to be added to the first set of data and causes the results not selected by the user but ranked higher than the selected result to be automatically added to the second set of data, the first and second sets of data persisted to a computer-readable storage medium.
2. The system of claim 1, the criteria comprising one or more of a document property, a context parameter, or a configuration.
3. The system of claim 2, the document property comprising one or more of a term that appears on a web page, a property of a Uniform Resource Locator (URL) identifying the web page, a property of a plurality of URLs that link to the web page, a property of a plurality of web pages that link to the web page, or a layout.
4. The system of claim 2, the context parameter comprising one of a word probability or a probability distribution
5. The system of claim 1, the tuning component is provided with training data to learn what properties of a document are indicative of the document being relevant to a user executing a search query from the entry point.

6. The system of claim 1, the tuning component configured to differentiate between a query result that is relevant to a search query context for a group of users and a query result that is non-relevant to the search query context for the group of users.
7. The system of claim 1, the tuning component employs statistical analysis in connection with filtering the search query results.
8. The system of claim 1, the tuning component generates one or more context parameters for a received query result, and compares the generated context parameters with a relevant context parameter and a non-relevant context parameter to determine whether the query result is relevant.
9. The system of claim 1, the tuning component further ranks the query results.
10. The system of claim 9, the ranking determined by the degree of relevance of the query result to the relevant data set and the non-relevant data set, the relevance is determined *via* one of a similarity measure or a confidence interval.
11. The system of claim 9, the ranking order comprising one of ascending or descending, from the most relevant result to the least relevant result.
12. The system of claim 1, the tuning component configured for a plurality of entry points associated with one or more groups of users.

13. A system that tunes a general-purpose search engine, comprising:
a filter component that receives search query results of a general-purpose search engine and parses relevant and non-relevant results based on training data associated with the entry point that provides a link employed to traverse to the general-purpose search engine, the training data comprises a first set of data categorized as relevant to a search context of a user for the entry point and a second set of data categorized as non-relevant to the search context of the user; and
a ranking component that sorts the filtered results in accordance with the training data for presentation to a user, wherein a user clicking a link associated with a search result from the sorted results causes the result to be added to the first set of data and causes the results whose links were not clicked by the user but that are ranked higher than the clicked result to be automatically added to the second set of data, the first and second sets of data persisted to a computer-readable storage medium.
14. The system of claim 13, the filter component parses the results as a function of one or more of a document property, a context parameter, or a configuration associated with the entry point.
15. The system of claim 13, the filter component trained to differentiate between a relevant and a non-relevant result *via* the training data.
16. The method of claim 13, the second set of data categorized as non-relevant comprising random data unrelated to the search context of the user for the entry point.
17. The system of claim 13, the filter component employs statistical analysis to determine whether a result is relevant or non-relevant to the entry point.
18. The system of claim 13, the ranking component employs a technique to determine the degree of relevance of the query results with respect to the relevant data set and the non-relevant data set.

19. The system of claim 18, the technique comprising one of a similarity measure or a confidence interval.
20. The system of claim 13, the ranking order comprising one of ascending and descending, from the most relevant result to the least relevant result.
21. The system of claim 18, the ranking performed on the relevant query results, the non-relevant results are discarded.
22. A method to filter and rank general-purpose search engine results based on criteria associated with an entry point, comprising:
- executing a query search with the general-purpose search engine accessed through a link associated with the entry point;
 - filtering the general-purpose search engine results by tuning the general-purpose search engine based on a set of training data associated with the entry point employed to access the general purpose search engine;
 - ranking the filtered general-purpose search engine results;
 - automatically storing a first query result selected by a user in a first data set categorized as relevant;
 - automatically storing at least one non-selected query result that is ranked higher than the first query result in a second data set categorized as non-relevant upon selection of the first query result; and
 - including the first data set and second data set in the set of training data associated with the entry point employed to access the general purpose search engine.
23. The method of claim 22, further comprising employing a statistical hypothesis to determine whether a result is relevant or non-relevant to a search context of the entry point.

24. The method of claim 23, the statistical hypothesis employing a threshold in connection with a probability distribution for relevant data and a probability distribution for non-relevant data, respective word probabilities are generated for the search query results and compared to the threshold, the probability distribution for relevant data and the probability distribution for non-relevant data to determine whether the results are relevant or non-relevant.

25. The method of claim 24, the threshold employed to bias the decision to mitigate one of a result being deemed non-relevant when the result is relevant or a result being deemed relevant when the result is non-relevant.

26. The method of claim 22, further employing a probability distribution analysis or machine learning in connection with the filtering and ranking, wherein suitable probability distributions include a Bernoulli, a binomial, a Pascal, a Poisson, an arcsine, a beta, a Cauchy, a chi-square with N degrees of freedom, an Erlang, a uniform, an exponential, a gamma, a Gaussian-univariate, a Gaussian-bivariate, a Laplace, a log-normal, a rice, a Weibull and a Rayleigh distribution, and the machine learning can classify based on one or more of a word occurrence, a distribution, a page layout, an inlink, and an outlink.

27. The method of claim 22, further comprising employing a statistical analysis to rank search query results.

28. The method of claim 27, the ranking comprising one of generating word probabilities and employing a confidence interval to determine relevance, and generating a similarity measure comprising one of a cosine distance, the Jaccard coefficient, an entropy-based measure, a divergence measure and/or a relative separation measure to determine similarity.

29. A method to customize a general-purpose search engine to improve context search query results, comprising:

tuning a general-purpose search engine for an entry point by employing a method further comprising:

providing a first set of data categorized as relevant that is used by a component to discern query results relevant to a search context of a user employing the entry point, the entry point provides a link employed to access the general-purpose search engine;

providing a second set of data categorized as non-relevant that is used by the component to discern query results unrelated to the search context, the first set of data and the second set of data are manually provided;

determining whether a query result is relevant or non-relevant to the search context based on the first set of relevant data and the second set of non-relevant data, each query result is compared with both the first set of data and second set of data to determine the relevance of the query result;

executing a search query with the general purpose search engine to obtain a ranked list of query results;

selecting a link associated with a query result from the list;

automatically adding the selected query result to the first set of data; and

automatically adding non-selected results from the list that are ranked higher than the selected query result to the second set of data upon selection of the selected query result.

30. The method of claim 29, the first set of data categorized as relevant comprising data associated with the search context of the user for the entry point.

31. The method of claim 29, the second set data categorized as non-relevant comprising random data unrelated to the search context of the user for the entry point.

32. The method of claim 29, further comprising providing information to associate respective query results with the entry point.

33. The method of claim 29, the first set of data categorized as relevant and the second set of data categorized as non-relevant employed to train the component to learn features that differentiate relevant data from non-relevant data.
34. A method to automatically customize a general-purpose search engine for an entry point, comprising:
- identifying the entry point;
 - executing a query search *via* the entry point that includes a link employed to route to the general-purpose search engine;
 - recording a first query result from a ranked list of query results returned from the executed query as relevant when a user views the document associated with the first query result;
 - recording at least one second query result whose associated document was not viewed by the user but that is ranked higher than the first query result as non-relevant when the first result is selected for viewing by the user; and
 - providing the recorded results to automatically train the filter for the entry point, in order to discriminate between results relevant to a search context of the user for the entry point and results non-relevant to the search context.
35. The method of claim 34, the set of relevant data comprising data associated with the search context of the user for the entry point.
36. The method of claim 34, the set of non-relevant data comprising data unrelated to the search context of the user for the entry point.
37. The method of claim 34, further comprising providing information to associate respective query results with the entry point.
38. The method of claim 34, the set of relevant data and the set of non-relevant data employed to train the component to learn the features that differentiate relevant data from non-relevant data.

39. The method of claim 34, the query results selected *via* a click thru technique employing a mouse to select a link associated with the query result by clicking on the link.
40. The method of claim 34, further comprising generating a word probability distribution for the relevant recorded results and a word probability distribution for the non-relevant recorded results.
41. (Cancelled).
42. A computer readable storage medium storing computer executable components that tunes a general-purpose search engine to improve context search query results, comprising:
- a component that receives search query results of a general-purpose search engine and filters the results based on training data sets associated with the search context of a user depending on the entry point that provides a link utilized to arrive at the general-purpose search engine, the training data sets include at least a first category of data explicitly defined to be relevant to the search context and a second category of data explicitly defined to be non-relevant to the search context; and
 - a component that ranks the filtered general-purpose search engine results according to the similarity of the search engine results to the training data sets, wherein selecting a link associated with a first search result from the ranked results causes the first result to be added to the first set of data and causes results that are ranked higher than the first result and have not been selected by the user to be automatically added to the second set of data.

43. A system that receives, filters and ranks general-purpose search engine results, comprising:

means for filtering general-purpose search engine results by determining whether a query result is relevant to a search context of a group of users, the search context is associated with an entry point that includes a link employed to navigate to the general-purpose search engine, the search context further having an associated first set of training data categorized as relevant to the context and an associated second set of training data categorized as non-relevant to the context; and

means for ranking the filtered general-purpose search engine results based on a relevance of the general-purpose search engine results to the search context of the group of users and the entry point as determined by a comparison of the search engine results with the first and second sets of training data, wherein a user viewing a document associated with a first search result from the ranked results causes the first result to be added to the first set of training data and causes the results that are unviewed but ranked higher than the first result to be automatically added to the second set of training data, the first and second sets of training data stored on a computer-readable storage medium.

IX. Evidence Appendix (37 C.F.R. §41.37(c)(1)(ix))

None.

X. Related Proceedings Appendix (37 C.F.R. §41.37(c)(1)(x))

None.